

University of Groningen

A Comparison between Young Students with and without Special Needs on Their Understanding of Scientific Concepts

van der Steen, Steffie; Steenbeek, Henderien; Wielinski, Janna; van Geert, Paul

Published in:
Education Research International

DOI:
[10.1155/2012/260403](https://doi.org/10.1155/2012/260403)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2012

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van der Steen, S., Steenbeek, H., Wielinski, J., & van Geert, P. (2012). A Comparison between Young Students with and without Special Needs on Their Understanding of Scientific Concepts. *Education Research International*, 2012, [260403]. <https://doi.org/10.1155/2012/260403>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Research Article

A Comparison between Young Students with and without Special Needs on Their Understanding of Scientific Concepts

Steffie Van Der Steen, Henderien Steenbeek, Janna Wielinski, and Paul Van Geert

Department of Developmental Psychology, Faculty of Behavioural and Social Sciences, University of Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, The Netherlands

Correspondence should be addressed to Steffie Van Der Steen, s.van.der.steen@rug.nl

Received 9 July 2012; Revised 20 September 2012; Accepted 25 October 2012

Academic Editor: David Neumann

Copyright © 2012 Steffie Van Der Steen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper examines whether young special needs (SN) students with emotional/behavioral difficulties (age 3–5, $n = 14$) reach lower understanding levels than regular students (age 3–5, $n = 17$) while working on two scientific tasks under a condition of scaffolding (e.g., follow-up questions depending on students' levels of understanding). Understanding was measured microgenetically, per utterance, using a scale related to Skill Theory. Monte Carlo analyses showed that SN students gave more wrong and (the lowest) Level 1 (single sensorimotor set) answers than regular students and fewer answers on (higher) Level 3 (sensorimotor system). However, no difference was found in their mean understanding level and mean number of answers. Both groups also had a comparable number of answers on the highest levels (Levels 4 and 5; single representation and representational mapping). These results do not point to substantial differences in scientific understanding between SN and regular students, as earlier studies using standardized tests have pointed out, and highlight the important role of scaffolding students' understanding. Standardized tests do not seem to indicate the bandwidth of possible scores students show or give an indication of their optimal scores, whereas a gap exists between student's task performance under conditions of individual performance and performance under a condition of support.

1. A Comparison between Students with and without Emotional and Behavioral Disorders on Their Understanding of Scientific Concepts

Numerous studies have shown that students with special needs (SN students) do not reach the level of academic performance of regular students, since their behavioral or emotional problems interfere with their ability to use their cognitive skills at an optimal level [1–3]. The focus of these studies is primarily on academic achievement, measured with summative assessment methods or standardized tests. However, do we obtain a valid picture of the capabilities, skills, and talents of students if we measure these with standardized tests, mostly referring to specific domains such as arithmetic and spelling? Instead, research should also focus on other domains, measures, and conditions of performance

in order to identify skills, and capabilities that would otherwise be missed. This paper aims to contribute to this matter by examining 31 regular and SN students' understanding of scientific concepts by using a microgenetic design and an alternative method of measuring understanding. The students (age 3–5) explored two scientific tasks under a condition of optimal scaffolding, meaning that they were encouraged and assisted by an adult while working on the tasks. The aim of this study is to examine whether differences between SN and regular students will be revealed in the process of building their understanding of scientific concepts, under the guidance of an experienced adult who provides adaptive scaffolding.

1.1. Children's Understanding of Scientific Concepts. Children's understanding of scientific concepts develops from a very young age on [4]. Recently, researchers have argued

the importance of studying the development of young children's understanding of scientific concepts. Young children's cognitive skills in the domain of science are the foundations of later literacy in this area and assist children in developing their reasoning about complex relationships [5]. The degree of understanding scientific concepts reflects the level of scientific thinking skills children can use while working on a problem solving task. Scientific thinking skills can be defined as the skills needed for describing a problem-solving situation, for forming hypotheses, testing hypotheses, and explaining as well as evaluating outcomes [6–10]. In the last decades, children's understanding of various scientific concepts has been studied. These studies predominantly focused on specific outcomes of individual learning processes, such as pre- and posttest scores on questionnaires [11]. In order to study students' understanding of scientific concepts, it is important to look not only at their achievements under a condition of individual performance, but also—even more importantly—under a condition in which they are supported [7].

The concept of scaffolding [12] comprises the temporary support of a child's learning process by an adult or more capable peer. The support is only temporary, since it is gradually reduced when the child reaches higher levels of competence and is capable of independent problem-solving [13]. Scaffolding unfolds dynamically [14] in that it describes not only how a particular level of knowledge or skill in a student changes as a result of the scaffolding process, but also how the scaffolding shifts as a result of the change in the student's performance. Teacher and student are engaged in a mutual process, in which the level of the student influences the level of the scaffold (which should be ahead of the first), while the level of the scaffold influences the level of the student. Given this definition of scaffolding as a dynamic mechanism of coupled teaching-learning processes, optimal scaffolding implies a student's optimal understanding as well as optimal teaching at the same time.

Researchers have pointed out the existence of a gap between children's task performance under conditions of individual performance (also referred to as the functional level) and performance under a condition of support (known as the optimal level, see [15]). This dichotomy dates back to the work of Vygotsky [16]. The general idea behind this dichotomy is that children do not show a single competence level, but instead vary across a range of possible levels. With help and guidance under a condition of scaffolding, students show an increase in understanding (or an increase in certain capacities), compared to a condition in which they work without receiving support [15]. In educational testing, unfortunately, emphasis is put on the functional level, meaning that what a student can do alone (an exception are dynamic testing methods, in which repeated testing is alternated with specific forms of feedback). The problem with these standardized methods of individual testing is twofold. First, it does not give us an idea of the student's learning potential, meaning the levels the student can reach with support, which will soon be mastered individually. Second, student's difficulties that interfere with scoring optimally on these tests, such as problems with focusing attention, or

understanding the wording of questions, remain unnoticed. Hence, the scores of students with special needs might not only reflect their understanding of a particular concept, but also to a great extent the problems they encounter in an individual testing situation. Under a condition of scaffolding, a teacher (or researcher) can not only attend to the student's needs in a testing situation, but also observe the capabilities of the student when receiving adequate support.

In this study, students were presented with two scientific tasks, while a researcher provided a variety of scaffolding techniques depending on the student's needs. This condition of optimal scaffolding differs from a dynamic testing (or assessment) method, which aims to measure students' learning potential in a particular domain by testing repeatedly and giving feedback after each test [17, 18]. Even though dynamic testing methods are used to unravel the process of learning, they are generally standardized, meaning that the questions, the moments of feedback, and the types of feedback are defined beforehand. In our condition of optimal scaffolding, we tried to create a naturalistic context somewhat similar to science classes in primary schools. That is, adult and student were constantly talking and working on the task; there were no longlasting monologues, and they did not take turns in manipulating the task. Moreover, feedback was not given at fixed intervals, but continuously during the interaction, mostly in the form of follow-up questions adapted to the student's answer, such as “Can you explain that?” or “How do you think we should figure that out?”

1.2. Special Needs Students. The Organization for Economic Co-operation and Development (OECD) defines students with special educational needs as those students who require “additional public and/or private resources to support their education” [19]. Since this definition is quite broad, the OECD has defined three cross-national subcategories in which special needs students can be divided: students with disabilities (e.g., sensory, motor, or neurological disabilities), students with difficulties (e.g., emotional and/or behavioral difficulties that have a negative effect on learning), and students with disadvantages (e.g., disadvantages due to socioeconomic or linguistic factors). Depending on the country and the student's condition, students with special needs receive extra resources within regular educational facilities, or are placed in special classrooms or schools. In the current research project, we visited special needs students with emotional and/or behavioral difficulties who were enrolled in special educational facilities. Most of these students were officially diagnosed with ADHD or mild forms of autism spectrum disorders (ASD), such as pervasive developmental disorder-not otherwise specified (PDD-NOS). A literature search showed that SN students with difficulties usually perform below the level of regular students [20, 21] on academic achievement tests that are usually standardized. This leads to the question whether a condition of optimal scaffolding would yield the same results.

In general, children diagnosed with ADHD show inattention (e.g., difficulty staying focused, often distracted and unorganized), hyperactivity (e.g., motoric restlessness,

excessive talking), and impulsivity (e.g., cannot wait for his/her turn, doing before thinking) [22], which seem to impair their ability to learn [23]. Luo and Li [24] found that the memory capacity (including short-term and working memory) of children with ADHD was impaired compared to that of typically developing children. Moreover, studies examining the processing level of children and adults with ADHD indicated that they have deficits in higher-level processing [25] and that they use different brain areas to encode complex or low-salient stimuli [26].

Children diagnosed with ASD are impaired in initiating and sustaining appropriate social interactions (e.g., maintaining relationships, limited social or emotional reciprocity) and communication (e.g., stereotyped use of language, impaired Theory of Mind). In addition, they often show limited and repetitive behavioral patterns [22]. Barnes et al. [27] stated that ASD students are not able to learn as easily as regular students, since they do not make deliberate use of their (social) environment, even though their implicit learning processes seem to be intact. Studies on higher-level processing of children with ASD showed that they exhibit difficulties when higher-level language processing (the use of meaning and context of a word) is needed to encode information [28].

Many SN students with difficulties (in our sample as well as in the broader population) have a combined diagnosis, such as pervasive developmental disorder-not otherwise specified (PDD-NOS) with hyperactivity symptoms, or ADHD with symptoms of oppositional deviant disorder (ODD). While there are differences with regard to the specific difficulties that students with different diagnoses encounter in learning situations, they do resemble each other in that SN students with difficulties generally display significant academic delays across all placements (including all forms of special education and general education; for a meta-analysis, see [20]), which do not seem to improve over time.

1.3. Measuring Children's Understanding of Scientific Concepts. In this study, the levels of understanding were operationalized by using a scale related to the 10 levels of Skill Theory, developed by Fischer [29]. Skill Theory focuses on the complexity and variability of children's skills, which consist of actions, verbalizations, and thinking abilities and the way these are constructed [15, 29]. One of the most powerful characteristics of Skill Theory [29] is that it extracts complexity from content, resulting in a content-independent measure of understanding. Because of this content-independent nature, Skill Theory enables researchers to compare understandings across multiple time points, contexts, persons, and age ranges [15, 30–32].

According to Fischer [15, 29], development in a particular domain goes through 10 levels of skills, hierarchically grouped into three tiers, that develop between 3 months and adulthood. The first tier consists of sensorimotor skills: simple connections of perceptions to actions or utterances. For example, the child states that two syringes are attached to one another by a tube. Any statements or actions going beyond the observation of elements, or observable

mechanisms, fall in the second and third tiers. The second tier is constituted of representational skills, understandings that go beyond current simple perception-action couplings, but are still based on them. That is, the term representation refers to the coordination of several sensorimotor skills at the same time [29]. Within the context of the two connected syringes, for example, the child can predict what happens if one of the pistons is pushed in, without literally touching or manipulating the syringe. Nonetheless, what he or she predicts depends not only on the context, but also on the sensorimotor skills mastered before. The third tier consists of abstractions, general rules that also apply to other situations. This would be an explanation about the relationship between pressure and volume inside a syringe [32]. Earlier (basic) skills form the basis of the more advanced skills across all tiers, that is, they are the building blocks of the higher levels.

Within each tier, sensorimotor, representational or abstract, three levels can be distinguished, each one is more complex than the previous one. The first one can be characterized as a single set, (e.g., a single representation or a single abstraction). The second level is a relation between two of these sets, which is referred to as a mapping. The third level is a system of sets, which is a relation between two mappings, in which each mapping consists of a relation between single sets. After this level, a new tier starts, which is divided in single sets, mappings, and systems as well [15].

Fischer and colleagues [15, 29, 32–34] showed that Skill Theory can not only describe and explain the development of skills on the long term, but also describe the microgenesis of problem solving [34]. When facing a new task or problem, even highly skilled adults go through the same cycles of skills. At the beginning they show skill levels that are mostly sensorimotor, which later build up to more elaborate levels. During a task, people do not go through the skill cycles in an orderly linear fashion. Instead, they repeatedly build up skill levels and regress before they obtain their highest possible level [33]. This variation between their highest and lowest possible complexity levels is also known as the developmental range. The highest levels within this range (reflecting the student's optimal level) are only reachable when the environment provides sufficient support [15, 33].

Given that students constantly vary within their developmental range (and given that we used a condition in which scaffolding was provided), it is important to measure understanding repeatedly during a task and capture the full range of skills students master in this context. Measuring students' understanding in a microgenetical way enables us to closely examine variations in students' understanding which reflect their thinking processes and prevents us from losing that information if we were measuring understanding at one point in time [35]. We therefore decided to register the skill theory levels of all task-related utterances. By looking not only at students' mean understanding level, but also at the distribution of their understanding levels, a more complete picture of their understanding can be revealed.

1.4. Research Questions and Hypotheses of This Study. This paper addresses the following questions. First, on average,

do the SN students reach a lower (Skill Theory) level of understanding than the regular students during the two scientific tasks while they are scaffolded by an adult? Second, if we look at the data from a more microgenetic point of view, does the proportion of the answer levels of SN students differ from that of the regular students during the scientific tasks?

To see whether the SN students would benefit from our scaffolding approach, we decided to take a falsification approach. If the scaffolding would *not* have a positive effect, we would, based on previous literature, expect to find that SN students' difficulties would impair them in crucial aspects relevant for the tasks, such as staying focused and being able to process complex information. In line with this, we would expect that (a_1) their mean level of understanding would be lower than that of the regular students, and that (a_2) they would have a lower mean number of correct task-related utterances (answers to questions), but (a_3) a higher mean number of incorrect task-related utterances (wrong answers to questions, i.e., mistakes). This leads to the hypothesis that (b_1) SN students would have a higher proportion of Level 1 (single sensorimotor set) and Level 2 (sensorimotor mapping) correct answers, which are the lowest Skill Theory levels. In contrast, regular students were expected (b_2) to answer more questions correctly on the three higher levels: Level 3 (sensorimotor system), Level 4 (single representation), and Level 5 (representational mapping). (We did not include levels higher than 5 into our hypotheses, because the ages associated with the emergence of these levels are above the age range of the students included in our study (see [15] for the ages of emergence).) However, if SN students would benefit from the scaffolding condition, we should be able to reject all hypotheses mentioned above and find no substantial differences between the two groups.

2. Method

2.1. Participants. The participants consisted of 14 Dutch SN students with emotional/behavioral difficulties (12 male, 2 female) enrolled in special educational facilities, and 17 Dutch regular students (10 male, 7 female) enrolled in regular educational facilities. Each group consisted of three cohorts recruited at the start of the study: 3-year olds ($M_{\text{age}} = 40$ months, $SD = 3.74$), 4-year olds ($M_{\text{age}} = 54$ months, $SD = 4.09$), and 5-year olds ($M_{\text{age}} = 65$ months, $SD = 4.52$). Although technically the 3-year-old students should be classified as preschoolers, we refer to them as students for the sake of simplicity. The two oldest SN cohorts ($n = 10$) attended kindergarten at a special needs primary school, and the youngest SN cohort ($n = 4$) attended a special needs daycare center. The two oldest ND cohorts ($n = 10$) attended kindergarten at a normal primary school, and the youngest ND cohort ($n = 7$) attended a regular daycare center. Recruitment took place at two schools and daycare centers in The Netherlands. Within these schools and centers, students' parents were asked if their children could participate in a study on scientific reasoning. All students whose parents responded positively were included in the study.

The SN students included in this study had emotional and/or behavioral difficulties that have a negative impact on their learning. They were officially diagnosed by psychological institutes or pedagogic professionals, most of them with ADHD (about 70% of the SN students), or a form of ASD (30% of the SN students). In The Netherlands, an official diagnosis is required to be able to enroll in a special school or educational facility. Given the severity of their problems and their developmental delays, these students were unable to follow the educational program offered at regular schools. The educational program in their special schools takes a slower pace and focuses more on the students' behavior and basic skills and knowledge. The lower percentage of female SN students (21.4%) is comparable to that of other mixed-gender studies on SN students with difficulties. Within the 13 mixed-gender studies included in their meta-analysis, Reid et al. [20] found percentages of females ranging from 9.3% to 63%, with an average percentage of 22.6%.

2.2. Procedure. During each visit, the students explored two scientific tasks individually, guided by a researcher, who was extensively trained into working with an adaptive protocol (see below). The first task involved the scientific concepts air pressure and Boyle's law, demonstrated by a task in which two syringes were attached to each other through a tube. When the piston of one syringe was pushed in, air travelled through the tube to the other syringe, which piston got pushed out as a consequence. During this task, syringes of different volumes were used. The second task during this visit was about the scientific concepts gravitation, inertia, and acceleration, which were demonstrated with a ball-run. Balls of different textures and weights were released at one end of the run, and slid down a path with different colors in order to determine which ball would come the farthest. The concepts of air pressure and gravity/inertia/acceleration were chosen because they provided a domain that was both limited and rich enough to study students' understanding of scientific concepts. Moreover, given their young age, the students had probably never encountered tasks like this, which meant that a continuous interaction with some form of scaffolding could be established.

To create a condition of optimal scaffolding, but also reach an acceptable level of standardization, an adaptive protocol was constructed. This guaranteed that all students were asked the basic questions that reflected the core building blocks of the scientific concepts incorporated in the task. At the same time, the protocol left enough space for students to show their understanding spontaneously and for the researcher to provide scaffolding when needed, without prompting the student with answers. This was done by asking follow-up questions related to the student's earlier answers encouraging the student to elaborate on an answer or asking for short explanations. For each task, the researcher showed the student the material and asked the student for its purpose and functioning at the very beginning. Afterwards—regardless whether the student answered the previous questions right, wrong, or at all—the student was encouraged to explore the material by him/herself. Subsequently,

the researcher asked questions about the task's functioning, as well as the underlying mechanisms, such as "Why does the piston of the other syringe get pushed out when you push the piston of this syringe?" The researcher gave the student time to answer, asked follow-up questions (related to the level of understanding as shown by the student), and encouraged him/her to think about the task and try out his/her ideas using the material. Even though students' answers were challenged sometimes, the feedback never included statements indicating whether the student was right or wrong. When the student could not give an explanation, the researcher proceeded with another question or subject. Each task took approximately 15 minutes. All interactions were recorded on video.

2.3. Coding of Verbal Understanding. In order to determine students' levels of understanding throughout the tasks, their verbal utterances were coded in four steps using the computer program MediaCoder [36]. The videos were coded in great detail, which enabled us to assign a range of understanding levels during a task. The first step in the coding procedure was the determination of the exact points in time when episodes of utterances started and ended. The second step involved the classification of all utterances of the student into several categories: descriptive, predictive, and explanatory answers/utterances; requests; content-related questions; other utterances. After this initial classification, meaningful units of the student's coherent utterances were formed in the third step of the coding procedure (units of analysis). This meant that the student's utterances about a single topic were combined. The unit of analysis ended when the next utterance of the student fell into another category, or when the researcher interrupted the student (e.g., by asking another question). However, if the researcher simply encouraged the student to tell more about the same topic, the unit of analysis would not end.

Lastly, the level of understanding per unit was determined by rating each unit on a ten level scale, which follows the model of Skill Theory [29]. These were the levels ranging from single sensorimotor sets (Level 1) to representational mappings (Level 5). At Level 1, students stated single characteristics of the task, such as "This ball is fast." At Level 2 (sensorimotor mapping), single characteristics were linked and comparisons between task elements were made, such as "This ball rolls faster than the other one." At Level 3 (sensorimotor system) students described aspects of the tasks in terms of causal observational relationships, such as "If I push the piston of this syringe, then the piston of the other one moves." At Level 4 (single representation), students were able to predict nonobservable characteristics and relations by saying for example, "I think this ball will come further than the other," or "Air causes the piston of the syringe to move." Lastly, at Level 5 (representational mapping), students could explain and predict in terms of two causal relationships including an additional step, for example, "The piston pushes the air, which travels through the tube to the other piston, which then gets pushed out by the air." Next to these five levels, an answer could also be

classified as a "mistake" when it was simply wrong, irrelevant, or when the student indicated that he or she did not know the answer to a question.

Videos were coded by two independent raters using a standardized coding book. For each round of coding (categories, units, and understanding levels), raters went through a training of coding three 15-minute video fragments and compared their codings with those of an expert-rater—the researcher who constructed the codebook. Initial differences between the raters and the expert-rater were solved through discussion. The codings of the third fragment were compared to the codings of the expert-rater and a percentage of agreement was calculated. The percentages of agreement on the third fragment were categories: 93% ($P < .01$), units: 94% ($P < .01$), and level of understanding: 92% ($P < .01$). The advantage of reporting simple percentages is that these are intuitively clear measures of agreement. Nevertheless, percentages provide no indication to what extent they depend on chance, which is why a P value (within brackets) was added [37]. The P values were calculated using a Monte Carlo procedure; for a description of this statistical procedure see Section 2.4.

2.4. Data Analysis. After coding SN and regular students' answers during both tasks, the frequencies for each level of understanding were determined. The mean level of understanding, the number of mistakes and answers, as well as the proportion of answers on each level were compared. For these comparisons, we used Monte Carlo permutation tests [38], which have great explanatory value in the case of small or skewed samples and result in reliable P values, since they do not assume any underlying distribution, or a minimum sample size [39]. Given our small sample size and skewed distribution of data, an ANOVA design (with accompanying assumptions) would decrease statistical power [40]. The Monte Carlo procedure estimates the probability that a certain difference between two groups is caused by chance alone. This is done by drawing a number of random samples from the original data (for this study 5000 random samples were drawn for each test), and determining how often the observed, or a bigger difference occurs in these random samples (positive cases). This number of positive cases is divided by the number of random samples in order to produce a P value for the tested difference, comprising the probability that the observed difference occurs in the distribution of 5000 random samples of the data. If the probability that this occurs is small, we can conclude that the observed difference is not merely caused by chance and, thus, that it is a legitimate difference.

Since we compared a number of differences between conditions and variables, we have decided to discuss only the interesting differences, which we defined as all differences for which the P value was equal to or smaller than .1 (which would support the hypotheses, and literature on academic differences between regular and SN students), and all differences that were contrary to our expectations (i.e., those results that would make us reject the hypotheses that the two groups differ, which would possibly indicate

TABLE 1: Mean, standard deviation, and minimum and maximum levels of understanding per group of students (regular and SN) and cohort.

Group	Age	N	Mean	SD	Min	Max
Regular students	All	17	2.54	0.27	0	5
	3	7	2.37	0.21	0	4
	4	5	2.46	0.21	0	4
	5	5	2.84	0.15	0	5
SN students	All	14	2.50	0.32	0	5
	3	4	2.38	0.27	0	4
	4	5	2.37	0.38	0	5
	5	5	2.74	0.17	0	5

the positive effect of scaffolding). The effect sizes of these differences (d) were calculated by dividing the difference in means by the standard deviation of the youngest age group (in case of within-group differences), or the standard deviation of the regular students (in case of between-group differences). These standard deviations were chosen because they were usually the biggest and, hence, yielded the most conservative measure of the effect size.

3. Results

3.1. Mean Levels of Understanding. Before testing our hypotheses, we first looked at the within-group differences in mean understanding level to see if similar patterns would evolve within each group. The results of the analysis are displayed in Table 1 and Figure 1. For the regular students, a significant difference in mean level of understanding was found between the 4-year olds and the 5-year olds, and between the 3- and 5-year olds ($P < .01$ for both differences, $d = 1.81$ and $d = 2.24$, resp.). For the SN students, a very similar pattern emerged: The 3-year olds and the 4-year olds differed significantly in their mean level of understanding from the 5-year olds ($P < .05$; $d = .97$ and $d = 1.33$, resp.).

3.1.1. Hypothesis a_1 : The Mean Level of Understanding Is Lower for the SN Students. Table 1 also shows the overall mean understanding level of the regular and SN students. Contrary to the hypothesis (a_1), the regular group reached only a slightly higher mean level of understanding ($M = 2.54$, $SD = .27$) compared to the SN group ($M = 2.50$, $SD = .32$). This difference was not statistically significant ($P = .36$). When looking at the differences in means for each age group, the results were similar. Even though the SN students had lower mean understanding levels in the two oldest age groups, and a comparable level of understanding in the youngest age group (see Figure 1), the differences with the regular students were too small to be statistically significant. We can therefore reject hypothesis a_1 and conclude that there are no significant differences in mean level of understanding, both in the group as a whole and across all age groups.

3.2. Mean Number of Correct Answers and Mean Number of Mistakes. Subsequently, the mean numbers of answers and mistakes were analyzed (see Table 2 and Figure 2). Again, the

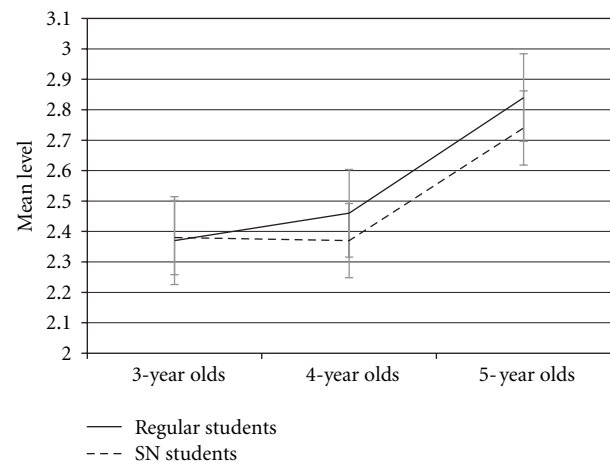


FIGURE 1: Mean understanding level (Y-axis) displayed by age X-axis) for each group.

within-group differences were explored first to see if we could detect similar patterns in the two groups. Within the regular group, the mean number of answers first decreased with age and then slightly increased, albeit not statistically significant. However, there were some significant differences regarding the mean number of mistakes for the regular group, that is, the difference between the 3- and 4-year olds ($P = .05$, $d = .77$), and the difference between the 3- and 5-year olds ($P < .05$, $d = .91$). The SN group showed a nonsignificant decrease in the mean number of answers between the 3- and the 4-year olds, and a significant increase between the 4- and the 5-year olds ($P < .05$, $d = 1.26$). Their mean number of mistakes, however, differed only slightly, and none of the differences between the age groups was statistically significant.

3.2.1. Hypotheses a_2 and a_3 : SN Students Have a Lower Mean Number of Correct Answers and a Higher Mean Number of Mistakes. The mean number of answers did not differ significantly ($P = .42$) between the two groups, which was in contrast with the hypothesis (a_2) that the mean number of answers would be lower in the SN group. The mean number of mistakes, however, was significantly higher for the SN students ($P < .01$, $d = .91$), which supported hypothesis a_3 . This was also found when we corrected for

TABLE 2: Means and standard deviations of the number of answers and mistakes per group of students (regular and SN) and cohort.

Group	Age	N	Mean answers	SD	Mean mistakes	SD
Regular students	All	17	116.59	40.95	37.35	19.44
	3	7	125.29	56.59	48.86	23.31
	4	5	109.6	34.33	31.0	11.79
	5	5	111.4	22.35	27.6	12.58
SN students	All	14	119.07	24.11	55.07	16.3
	3	4	117.25	30.08	60.0	25.13
	4	5	107.0	20.35	55.6	15.08
	5	5	132.6	19.55	50.6	10.64

the number of answers, that is, when we compared the mistakes proportional to the total number of answers, which yielded a higher proportion (0.46) for the SN students compared to the proportion (0.32) for the regular students ($P < .01$, $d = 1.45$).

When looking at the different age groups, the 3-year-old regular students did not differ significantly from the 3-year-old SN students in terms of their mean number of answers, but also not in their mean number of mistakes. However, the ratio wrong/total number of answers of the 3-year old SN students (0.5) was significantly higher than that of the 3-year-old regular students (0.39), $P < .05$, $d = 1.19$. The mean number of answers of the 4-year-old regular students also did not differ from that of the SN students. That said, their mean number of mistakes was significantly higher ($P = .01$, $d = 2.09$). This was also the case when the ratio wrong/total number of answers was compared. The ratio of the 4-year old SN students was significantly higher (0.52) than that of the regular students (0.29), $P < .01$, $d = 3.47$. Lastly, the 5-year-old regular and SN students differed significantly with respect to both their mean number of answers and their mean number of mistakes ($P = .05$, $d = .95$ and $P < .01$, $d = 1.83$, resp.). Note that the 5-year-old SN students answered more questions than the regular students ($M = 132.6$, $SD = 19.55$ versus $M = 111.4$, $SD = 22.35$), contrary to hypothesis a_2 . Nevertheless, they also made more mistakes ($M = 50.6$, $SD = 10.64$ versus $M = 27.6$, $SD = 12.58$), and the ratio wrong/total number of answers was higher for the SN students than for the regular students (0.38 and 0.24 resp., $P < .01$, $d = 1.95$), which was in line with what was expected (a_3).

To summarize, we found no evidence for the hypothesis that SN students have a lower mean number of correct answers across all age groups, so we can reject hypothesis a_2 . On the other hand, we did find evidence for the hypothesis that SN students have a higher mean number of mistakes, and cannot reject hypothesis a_3 .

3.3. The Proportion of the (Skill Theory) Answer Levels. In order to answer whether the distribution of the answer levels of SN students differed from that of the regular students, the number of answers was counted for each level and divided by the total number of answers within each (age) group. To test the differences between the groups, the mean proportions were used (see Table 3).

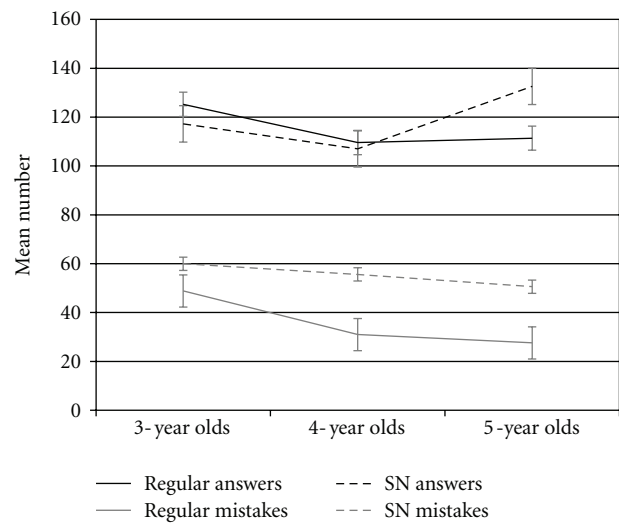


FIGURE 2: Mean numbers (Y-axis) of answers and mistakes by age (X-axis) for each group.

3.3.1. Hypothesis b_1 : SN Students Have a Higher Proportion of Correct Answers on Levels 1 and 2. When we compared the regular students with the SN students across all age groups (see the left upper graph of Figure 3), SN students had a significantly higher proportion of Level 1 answers ($P < .01$, $d = 2.0$) as was hypothesized. However, the regular group had more answers on Level 2 ($P = .05$, $d = .55$), which was in contrast with hypothesis b_1 . When looking at the 3-year olds, a similar difference between the groups emerged for Level 1 ($P < .05$, $d = 1.06$). The 4-year-old SN students also had a higher proportion of Level 1 answers compared to their regular peers ($P < .01$, $d = 4.4$), and given the large effect size, this seems to be a considerable difference. The 4-year-old regular students had a higher mean proportion of level 2 answers than the SN students ($P = .05$, $d = 1.06$), which was in contrast with hypothesis b_1 . For the 5-year-old students, the difference in the proportion of Level 1 answers between the SN students and the regular students was significant ($P < .01$, $d = 3.3$). In sum, SN students had indeed a higher proportion of correct Level 1 answers across all age groups, which was in line with hypothesis b_1 . For Level 2 answers, however, the overall group of regular students had a significantly higher proportion, as well as the 4-year olds.

TABLE 3: Proportions of correct answers per level of understanding (the number of correct answers for each level divided by the total number of correct answers of each (age) group).

Group	Age	N	1	2	3	4	5
Regular students	All	17	0.04	0.60	0.14	0.21	0.005
	3	7	0.05	0.67	0.13	0.17	0.0
	4	5	0.04	0.65	0.12	0.19	0.0
	5	5	0.02	0.47	0.17	0.32	0.01
SN students	All	14	0.13	0.51	0.09	0.26	0.007
	3	4	0.11	0.64	0.02	0.23	0.0
	4	5	0.18	0.51	0.09	0.21	0.01
	5	5	0.09	0.42	0.15	0.33	0.01

For the 3- and 5-year olds, no significant difference in the proportion of Level 2 answers was found. Hence, the results for the proportion of Level 2 answers are not in line with hypothesis b_1 .

3.3.2. Hypothesis b_2 : Regular Students Have a Higher Proportion of Correct Answers on Levels 3, 4, and 5. In the overall group, the regular students had a higher proportion of Level 3 answers ($P = .06$, $d = .49$), which supported hypothesis b_2 . On Level 4, however, the SN students outperformed the regular students, which was unexpected ($P = .1$, $d = .49$). No significant difference between the groups was found for Level 5 ($P = .31$). When looking at the separate age groups, the 3-year olds showed a similar difference between regular and SN students on Level 3 ($P < 0.05$, $d = .86$). For this age group, the difference on Level 4 was also noteworthy, since the 3-year-old SN students had a higher proportion of answers on this level than the regular students ($P = .07$, $d = 1.04$). For the 4- and 5-year olds, the differences between the groups on Level 3, 4 and 5 were too small to be statistically significant. To conclude, the only evidence in line with hypothesis b_2 was found for the proportion of Level 3 answers in the overall group and for the 3-year olds. All other differences were not in line with hypothesis b_2 .

Figure 3 shows the proportion of answer levels, both for the groups as a whole and for the separate age groups. Despite some small differences (mostly on Levels 1 and 2), the shape of the graphs of the two groups is strikingly similar, with peaks at Levels 2 and 4, low values at Levels 1 and 5, and a dip at Level 3. In the graph of the 3-year olds (right upper graph), the dip at Level 3 is clearly lower for the SN students than for the regular students, whereas the rest of the proportions seem to be similar. The graphs for the 4- and 5- year-old students (lower two graphs) look even more similar. The difference in the proportion of Level 3 answers is smaller for these age groups, and the proportions of answers on Levels 4 and 5 seem to be equal.

4. Discussion

The aim of this research was to examine whether differences between 3- to 5-year-old SN and regular students would emerge in the process of building their understanding of scientific concepts while working on two scientific tasks: one

about air pressure and Boyle's law, and one about gravity, inertia, and acceleration, under a condition of optimal scaffolding in a natural setting.

4.1. Overview of Our Findings. With regard to the mean level of understanding, the hypotheses that SN students' mean level of understanding would be lower (a_1), and that they would have a lower mean number of answers (a_2) must be rejected. The hypothesis that SN students would make more mistakes (a_3) was the only hypothesis that was mostly supported by our data. That is, the overall SN group made more mistakes than the regular group. This was also the case when the 4- and 5-year-old SN and regular students were compared. For the 3-year olds, no difference was found when absolute measures were compared; however, the ratio wrong/total answers was significantly higher for the 3-year-old SN students.

In line with hypothesis b_1 , SN students had a higher proportion of Level 1 (single sensorimotor set) answers compared to the regular group. Contrary to this hypothesis, however, the regular students outperformed the SN students on Level 2 (sensorimotor mapping) in the overall group and most age groups. In addition, the regular students had indeed a higher proportion of Level 3 (sensorimotor system) answers (hypothesis b_2), but this was mostly caused by the difference between the 3-year-old SN and regular students. On Levels 4 and 5 (single representation and representational mapping), the groups scored roughly equal, which was not in line with hypothesis b_2 . In general, most findings were in contrast with the hypotheses and previous research.

4.2. The Positive Effects of Optimal Scaffolding Conditions. In the last years, studies showed that students with special needs are not learning the required basic academic skills and perform below the level of regular students across several domains. Most of these studies focused on math and reading skills [1–3], measured with standardized tests [20], although some have focused on scientific thinking [21]. The outcomes of these studies are in contrast with the performance of SN students under our optimal scaffolding condition. In fact, our results are even in contrast with the standardized test scores of the SN students included in this study, on which they performed below the regular students. Most Dutch schools take part in a national assessment program (Cito)

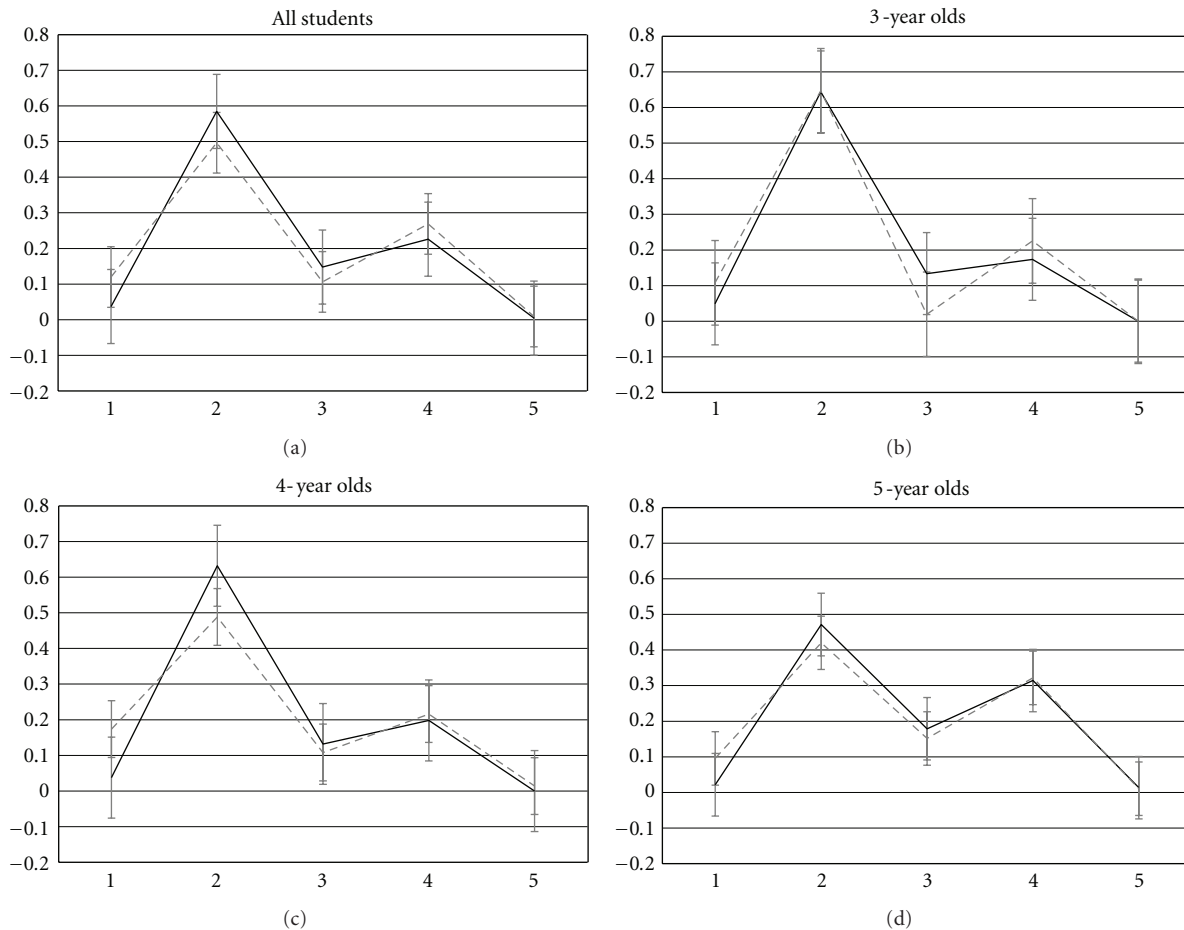


FIGURE 3: The proportion (Y-axis) of the answer levels 1–5 (X-axis) for all students, and the 3 age groups. Regular students are displayed by the solid line, SN students by the dashed line.

and regularly evaluate their students' progress on several subjects, such as math and language skills. We collected the regular and SN students' test scores on their first Cito language and math tests administered in kindergarten. On both tests, students could get a score from 1 (E, lowest score) to 5 (A, highest score). We obtained data for 28 of our students; the data of three SN students were not available, because they had not yet been tested. Taking the mean score of these two tests, our regular students had a score of 4.4 on average, whereas the SN students had a score of 3.68. Using a Monte Carlo test we found this difference to be statistically significant ($P < .05$), with an effect size (d) of .67. This means that at this time, the regular students performed two-third of a standard deviation better on these two academic tests compared to the SN students in our sample.

The question arises whether the skills and performances examined with standardized tests are similar to those in this research. Standardized tests do not indicate the bandwidth of possible scores children show or give an indication of their optimal scores, whereas researchers have pointed out the existence of a gap between children's task performance under conditions of individual performance and performance under a condition of support [15]. In other words, the context in which one assesses students' capabilities

influences the results to a great extent. This context can be a difference not only in terms of measurement setting or presentation of tasks (standardized versus scaffolding), but also in terms of the type and phrasing of questions. In a study of Ayoub et al. [41], maltreated children (42 months old) were not able to retell stories involving nice interactions as accurately as nonmaltreated children. However, both groups showed roughly the same scores when asked to re-tell stories involving mean interactions. The authors conclude that maltreated children are not cognitively impaired in the traditional sense, but instead have learned to focus more on negative aspects, which can be an adaptive response to threat.

The current research shows that special needs students with behavioral difficulties perform on the same level as regular students on tasks requiring scientific thinking and reasoning, if they are guided by an adult who uses appropriate scaffolding techniques to respond to the student's emotional and cognitive needs. On the other hand, standardized tests in math and language seem to be too demanding. Cooper et al. [42] indicated that standardized test scores are not always appropriate to measure problem-solving skills of SN students. In their study on problem-solving, which included experiential science materials, a mentoring component, and assessment of students' scientific

products instead of their test scores, the problem-solving skills of SN students were comparable to those of regular students. This study also seems to indicate that SN students' scientific problem solving skills (and their understanding, which reflects the level of these skills) are more advanced in conditions in which they receive adaptive support from the environment. Their individual performance, in the literature mostly measured by standardized tests (and in the case of our sample by math and language tests), might not accurately reflect the SN students' full potential.

4.3. Standardized Tests versus Conditions of Scaffolding: What Do They Measure? For many SN students, the validity of (standardized) tests depends on the accessibility of test items and tasks. As an example, a dyslexic student's score on a standardized math test might not only reflect the student's math skills, but also the ability to read the test items and instructions [43]. Hence, standardized tests do not only measure the constructs they claim, and students' test scores might reflect some construct-irrelevant noise. The students included in our study were not print-disabled, but had other difficulties, and formal testing situations might be unable to meet their individual needs. These needs might well be met in a scaffolding condition, in which the researcher continuously draws the student's attention, changes the wording of questions if necessary, and uses follow-up questions to get a complete picture of the student's understanding, or challenges an earlier given answer. Moreover, the hands-on tasks used in this study enabled the students to try out their ideas and, if necessary, change their explanations of the mechanisms at work.

Scaffolding does not mean that students get so much help that they simply surpass their own level of performance, nor does it mean that students are prompted with answers. Instead, scaffolding sets a context in which students can access the upper section of their range of possible scores. Although scaffolding is seldom used in summative assessment methods, Almond et al. [43] note that scaffolding provides students with supports that help them to answer questions at their individual level, which allows us to better measure students' knowledge and skills. Under a condition of scaffolding, teachers can see what students do know about a particular item, instead of simply marking their answer as wrong or incomplete. This study shows that when children are in a situation in which scaffolding is applied frequently, differences between special needs and regular children almost disappear. We therefore advise teachers in special educational settings to use a wide range of adaptive scaffolding techniques (follow-up questions, encouragement, instructions, and feedback) during their lessons. In doing so, teachers can pay particular attention to the mistakes SN students make (which they made more in this study compared to the regular students) and encourage them to elaborate on the correct parts of their thinking. By carefully watching students' responses in the classroom, the difficulties of SN students can be detected and further addressed by using scaffolding techniques. For example, the 3-year-old SN students in this study had difficulties in expressing causal

relationships, that is, they had significantly less answers on Level 3 (sensorimotor system). These young students might benefit from more scaffolding directed towards this type of reasoning.

New initiatives show that scaffolding conditions are not as far from formal testing situations as one would imagine. Research suggests that applying universal design principles can improve testing of SN students with difficulties, by providing alternative forms of instructions (e.g., not only text, but also graphs or pictures, or videos), alternative forms of expression (e.g., not only writing down answers, but also drawing or using graphic organizers), and alternative forms of engagement (e.g., choosing a topic for a test on reading comprehension) [43, 44].

4.4. Suggestions for Future Research. The number of SN students is growing [45], and therefore it becomes more and more important to assess not only their disabilities, but also their capabilities both in the academic context and beyond. Identifying their strengths and providing help to make use of these strengths could support students in developing a more positive self-concept and self-efficacy, which they often lack due to failure experiences in the academic context [42]. Future research should investigate what characteristics of students' environment (materials, tasks, and interactions with adults or peers) support the development of their (scientific thinking) skills, in order to advise teachers, parents and therapists regarding the optimal adjustment of academic contexts to students' individual needs. In addition, the microgenetic approach we used (coding per utterance) yielded a continuous measurement of students' understanding and showed that understanding shifts regularly between levels over time (see also [34]). Measuring understanding using aggregated data of single tests might prevent us from detecting these variations in students' understanding and could possibly lead to inaccurate measures. Further research should both investigate the benefits of scaffolding for SN students in more detail, as well as the variations in their academic achievements over time. The results of these studies can then be used to optimize standardized tests, so that SN students can make optimal use of these situations.

References

- [1] T. E. Scruggs and M. A. Mastropieri, "Academic characteristics of behaviorally disordered and learning disabled students," *Behavioral Disorders*, vol. 11, pp. 184–190, 1986.
- [2] M. H. Epstein, D. Kinder, and B. Bursuck, "The academic status of adolescents with behavioral disorders," *Behavioral Disorders*, vol. 14, pp. 157–165, 1989.
- [3] A. L. Trout, P. D. Nordness, C. D. Pierce, and M. H. Epstein, "Research on the academic status of children with emotional and behavioral disorders: a review of the literature from 1961 to 2000," *Journal of Emotional and Behavioral Disorders*, vol. 11, no. 4, pp. 198–210, 2003.
- [4] R. S. Siegler and M. W. Alibali, *Children's Thinking*, Pearson Prentice Hall, New York, NY, USA, 2005.

- [5] National Research Council, *Mathematical and Scientific Development in Early Childhood: A Workshop Summary*, The National Academies Press, Washington, DC, USA, 2005.
- [6] C. Zimmerman, "The development of scientific reasoning skills," *Developmental Review*, vol. 20, no. 1, pp. 99–149, 2000.
- [7] C. Zimmerman, "The development of scientific thinking skills in elementary and middle school," *Developmental Review*, vol. 27, no. 2, pp. 172–223, 2007.
- [8] B. Koslowski, L. Okagaki, C. Lorenz, and D. Umbach, "When covariation is not enough: the role of causal mechanism, sampling method, and sample size in causal reasoning," *Child Development*, vol. 60, no. 6, pp. 1316–1327, 1989.
- [9] D. Kuhn and S. Franklin, "The second decade: What develops (and how)," in *Handbook of Child Psychology: Vol. 2. Cognition, Perception and Language*, W. Damon, R. M. Lerner, D. Kuhn, and R. S. Siegler, Eds., pp. 953–993, John Wiley & Sons, Hoboken, NJ, USA, 6th edition, 2006.
- [10] F. Wilkening and B. Sodian, "Scientific reasoning in young children: introduction," *Swiss Journal of Psychology*, vol. 64, no. 3, pp. 137–139, 2005.
- [11] S. van der Steen, H. Steenbeek, and P. van Geert, "Using the dynamics of a person context system to describe children's understanding of air pressure," in *Current Topics in Children's Learning and Cognition*, H. Kloos, B. J. Morris, and J. L. Amaral, Eds., pp. 21–44, 2012, <http://www.intechopen.com/books/current-topics-in-children-s-learning-and-cognition/using-the-dynamics-of-a-person-context-system-to-describe-children-s-understanding-of-air-pressure>.
- [12] D. Wood, H. Wood, and D. Middleton, "An experimental evaluation of four face-to-face teaching strategies," *International Journal of Behavioral Development*, vol. 1, pp. 131–147, 1978.
- [13] M. Pressley, K. Hogan, R. Wharton-McDonald, and J. Mistretta, "The challenges of instructional scaffolding: The challenges of instruction that supports student thinking," *Learning Disabilities Research & Practice*, vol. 11, no. 3, pp. 138–146, 1996.
- [14] P. van Geert and H. Steenbeek, "The dynamics of scaffolding," *New Ideas in Psychology*, vol. 23, no. 3, pp. 115–128, 2005.
- [15] K. W. Fischer and T. R. Bidell, "Dynamic development of action, thought and emotion," in *Theoretical Models of Human Development. Handbook of Child Psychology*, R. M. Lerner and W. Damon, Eds., vol. 1, pp. 313–399, John Wiley & Sons, Hoboken, NJ, USA, 6th edition, 2006.
- [16] L. S. Vygotsky, *Thought and Language*, The MIT Press, Cambridge, Mass, USA, 1986, Original work published in 1934.
- [17] C. S. Lidz, *Practitioner's Guide to Dynamic Assessment*, Guilford Press, New York, NY, USA, 1991.
- [18] R. J. Sternberg and E. L. Grigorenko, *Dynamic Testing: The Nature and Measurement of Learning Potential*, Cambridge University Press, Cambridge, UK, 2002.
- [19] OECD, *Students With Disabilities, Learning Difficulties and Disadvantages: Statistics and Indicators*, OECD Publishing, 2005.
- [20] R. Reid, J. E. Gonzalez, P. D. Nordness, A. Trout, and M. H. Epstein, "A meta-analysis of the academic status of students with emotional/behavioral disturbance," *Journal of Special Education*, vol. 38, no. 3, pp. 130–143, 2004.
- [21] P. Mooney, M. H. Epstein, R. Reid, and J. R. Nelson, "Status of and trends in academic intervention research for students with emotional disturbance," *Remedial and Special Education*, vol. 24, no. 5, pp. 273–287, 2003.
- [22] American Psychological Association, *Diagnostic and Statistical Manual of Mental Disorders*, American Psychological Association, Washington, DC, USA, 4th edition, 2000.
- [23] T. Humphries, "Attention-deficit/hyperactivity disorder," in *A Comprehensive Guide to Intellectual and Developmental Disabilities*, I. Brown and M. Percy, Eds., pp. 295–307, Paul H Brookes Publishing, Baltimore, Md, USA, 2007.
- [24] X. Luo and X. Li, "Memory and learning ability of children with ADHD," *Chinese Mental Health Journal*, vol. 17, no. 3, pp. 188–190, 2003.
- [25] A. C. Kalff, L. M. J. de Sonnevill, P. P. M. Hurks et al., "Low- and high-level controlled processing in executive motor control tasks in 5–6-year-old children at risk of ADHD," *Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 44, no. 7, pp. 1049–1057, 2003.
- [26] T. Sigi Hale, S. Bookheimer, J. J. McGough, J. M. Phillips, and J. T. McCracken, "Atypical brain activation during simple & complex levels of processing in adult ADHD: an fMRI study," *Journal of Attention Disorders*, vol. 11, no. 2, pp. 125–140, 2007.
- [27] K. A. Barnes, J. H. Howard Jr., D. V. Howard et al., "Intact implicit learning of spatial context and temporal sequences in childhood autism spectrum disorder," *Neuropsychology*, vol. 22, no. 5, pp. 563–570, 2008.
- [28] I. L. J. Noens and I. A. V. Berckelaer-Onnes, "Captured by details: sense-making, language and communication in autism," *Journal of Communication Disorders*, vol. 38, no. 2, pp. 123–141, 2005.
- [29] K. W. Fischer, "A theory of cognitive development: the control and construction of hierarchies of skills," *Psychological Review*, vol. 87, no. 6, pp. 477–531, 1980.
- [30] K. W. Fischer and N. Granott, "Beyond one-dimensional change: parallel, concurrent, socially distributed processes in learning and development," *Human Development*, vol. 38, pp. 302–314, 1995.
- [31] K. W. Fischer and R. Corrigan, "A skill approach to language development," in *Language Behavior in Infancy and Early Childhood*, R. Stark, Ed., pp. 245–273, Elsevier, Amsterdam, The Netherlands, 1981.
- [32] M. Schwartz and K. W. Fischer, "Building general knowledge and skill: cognition and microdevelopment in science learning," in *Cognitive Developmental Change: Theories, Models, and Measurement*, A. Demetriou and A. Raftopoulos, Eds., pp. 157–185, Cambridge University Press, Cambridge, UK, 2004.
- [33] Z. Yan and K. W. Fischer, "Always under construction—dynamic variations in adult cognitive microdevelopment," *Human Development*, vol. 45, no. 3, pp. 141–160, 2002.
- [34] N. Granott and J. Parziale, "Microdevelopment: a process-oriented perspective for studying development and learning," in *Micro Development: Transition Processes in Development and Learning*, N. Granott and J. Parziale, Eds., pp. 1–28, Cambridge University Press, Cambridge, Mass, USA, 2002.
- [35] R. S. Siegler, "Microgenetic analyses of learning," in *Handbook of Child Psychology: Volume 2: Cognition, Perception, and Language*, W. Damon, R. M. Lerner, D. Kuhn, and R. S. Siegler, Eds., pp. 464–510, John Wiley & Sons, Hoboken, NJ, USA, 6th edition, 2006.
- [36] J. Bos and H. Steenbeek, *Mediacoder, A Simple Application For Coding Behavior within Media Files*, University of Groningen, Groningen, The Netherlands, 2007.
- [37] P. van Geert and M. van Dijk, "Ambiguity in child language: the problem of interobserver reliability in ambiguous observation data," *First Language*, vol. 23, no. 3, pp. 259–284, 2003.
- [38] J. B. Todman and P. Dugard, *Single-Case and Small-N Experimental Designs: A Practical Guide To Randomization Tests*, Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2001.
- [39] P. Van Geert, H. Steenbeek, and S. Kunnen, "Monte Carlo techniques: statistical simulation for developmental data," in

- A Dynamical Systems Approach to Adolescent Development*, S. Kunnen, Ed., pp. 43–53, Psychology Press, Hove, UK, 2012.
- [40] T. Baguley, “Can we be confident in our statistics?” *The Psychologist*, vol. 25, pp. 128–129, 2012.
 - [41] C. C. Ayoub, E. O’Connor, G. Rappolt-Schlichtmann et al., “Cognitive and emotional differences in young maltreated children: a translational application of dynamic skill theory,” *Development and Psychopathology*, vol. 18, no. 3, pp. 679–706, 2006.
 - [42] C. R. Cooper, S. M. Baum, and T. W. Neu, “Developing scientific talent in students with special needs. An alternative model for identification, curriculum, and assessment,” *Journal of Secondary Gifted Education*, vol. 15, no. 4, pp. 162–169, 2004.
 - [43] P. Almond, P. Winter, R. Cameto et al., “Technology-enabled and universally designed assessment: considering access in measuring the achievement of students with disabilities—a foundation for research,” *Journal of Technology, Learning, and Assessment*, vol. 10, no. 5, pp. 1–52, 2010.
 - [44] R. P. Dolan and T. E. Hall, “Universal design for learning: implications for large scale assessment,” *IDA Perspectives*, vol. 27, no. 4, pp. 22–25, 2001.
 - [45] U.S. Public Health Service, *Report of the Surgeon General’s Conference on Children’s Mental Health: A National Action Agenda*, Department of Health and Human Services, Washington, DC, USA, 1999, <http://www.surgeongeneral.gov/topics/cmh/childreport.html>.

